



Nigerian Journal of Technology (NIJOTECH)

Vol. 35, No. 1, January 2016, pp. 210 – 218

Copyright© Faculty of Engineering, University of Nigeria, Nsukka,

Print ISSN: 0331-8443, Electronic ISSN: 2467-8821

www.nijotech.com<http://dx.doi.org/10.4314/njt.v35i1.28>

EUDIOMETRIC THEORETIC-APPROACH TO MODELLING THE ASSIMILATIVE CAPACITY OF A RIVER: INCORPORATION OF BOOTSTRAPPING NEEDFUL FOR SENSITIVITY ANALYSIS

C. M. Chiejine^{1, *}, A. C. Igboanugo² and L. I. N. Ezemonye³

¹ DEPT OF ELECTRICAL/ELECTRONIC ENGINEERING, DELTA STATE POLYTECHNIC, OGWASHI-UKU, DELTA STATE, NIGERIA

² DEPARTMENT OF PRODUCTION ENGINEERING, UNIVERSITY OF BENIN, BENIN CITY, EDO STATE NIGERIA

³ DEPARTMENT OF ANIMAL AND ENVIRONMENTAL BIOLOGY, UNIVERSITY OF BENIN, BENIN CITY, EDO STATE .NIGERIA

E-mail addresses: ¹ pstchiejine@yahoo.com, ² dracigboanugo@yahoo.com, ³ ezemslaw@yahoo.com

ABSTRACT

The mathematical physics underlying the adsorption and subsequent desorption of dissolved oxygen (DO) in a water body subject to effluent loading had been rarely investigated. The current state of play in this field although reflects use of different analysis, the combine use of hat matrix and bootstrapping techniques to study the phenomenon of chemical adsorption and desorption of DO at molecular level in a polluted waterbody has not been thoroughly investigated. This study seeks to use a matrix projector, \hat{H} , to cast virtual spectral rays on pollutant loadings in a water body and in the process unravel the dynamics of chemical and biological gravitation of dissolved oxygen towards constituents of effluent pollutants in water body. This approach is anchored on the ordinary least squares methodology of multivariate linear regression. The method hypothesized is studied by a mathematical physics analysis of the phenomenon. Bootstrapping was used to establish means and variances of regression parameters, and subsequently, the confidence intervals of point estimates of parameters. Tricking technique adopted facilitated the development of extreme values of the dissolved oxygen and hence the supremum and infimum of assimilative capacity of the river which fluctuates with intensity of effluent loadings and season of the year (rainy, dry, and harmattan seasons). The result of bootstrapping revealed that assimilative capacity fluctuated widely from the values detected by point estimates of regression parameters thus suggesting that tricking of regression parameters, in turn, tunes up the regression model, and hence, fine tunes the value of assimilative capacity through necessary adjustments of model parameters. The results of this study obviates the need to deploy eudiometer for laborious direct measurement of dissolved oxygen in a body of polluted water. Thus an elegant technique for crossing the stream where it is shallowest has been developed in this study. The method is considered as a great improvement on previous approaches that seem to dawdle.

Keywords: eudiometry, bowl, assimilative capacity, bootstrapping, model tricking.

1. INTRODUCTION

The dynamics of interaction between pollutants from effluents and dissolved oxygen (DO) in a body of water has not been well studied. But it is known, however not in detail, that each pollutant in effluent discharge somehow react when mixed in water and in the process adsorb and subsequently desorb oxygen by virtue of relative valance between it and molecules of the DO in the body of water. And, as earlier stated, the mathematics underpinning the adsorption/desorption of oxygen molecule by each pollutant from

the body of water is not well understood [1]. This study attempts to use projector matrix \hat{H} , in the sense of ordinary least squares (OLS) of the multiple linear regression model, to unravel the phenomenon of dissolved oxygen adsorption and desorption in a body of water subject to effluent loading. The \hat{H} matrix is conceived as virtual spectral rays incident upon a medium of water in which pollutant are injected as a plume. The virtual spectral signal (VSS) is assumed to energize the molecules of the pollutants such that, based on their relative valance, oxygen molecules will

gravitate towards each pollutant according to their respective potentials. Again, gobs and gobs of different pollutants in the effluent plume is treated as a column vector of independent variables x_i , while the balance of DO in the medium which determines the assimilative capacity (AC), is treated as response or explanatory variable, y . The problem in this modelling methodology is to apply virtual eudiometry as black box to estimate the amount of DO at any time based on sample realization from the polluted water body. Thus, the chief parameter of interest from the sample realization is the $\hat{\beta}$ which can be expressed as a linear combination of the column vector: $[x_1, x_2, x_3, \dots, x_n]^T$ to give the estimate or ballpark of the population values of the parameter that estimates the DO (y). It is easy to visualize that estimate of the parameters are point values and we need a special technique of bootstrapping to determine mean and variance of each of the values of the point parameters so that the confidence intervals which facilitate tricking of the mother parameters can be determined.

Whereas oodles of studies have addressed the pollution in Ikpoba River using proximate analysis and related approaches, the balance of literature is still grossly deficient in the determination of assimilative capacity and mathematical physics analysis of pollutant dispersion in both far-field and near-field. See for example [1-4].

Again, few authors have applied linear regression in investigating effluent dispersion (see, for example also [5-8]. Apart from [9] which applied bootstrapping technique in a study to analyze the 'Sobol' sensitivity of a complex environmental model, very little have been researched using bootstrapping techniques to undertake sensitivity analysis on fluctuation of the DO in a body of polluted water. The current study therefore is a kind of effort to breach the perceived frontier of knowledge. The knowledge of the dynamics of relative adsorption of DO by the constituent pollutants, as this study affords, is important in the sense that it provides insight and enlightenment into better understanding of how the constituent pollutants contribute most to water body pollution and hence serve as a useful guide to modelling and corrective actions. Again, it will also help to test the robustness of the optimal solution.

The solution to the nagging problem of surface water pollution resulting from industrial effluents and municipal surface run-off require reliable and scientifically proven information. Such information

can only be made available through regular monitoring programmes. Regular monitoring usually result in large volume of complex data matrix which requires accurate interpretation and significant conclusions.

Relatedly, [10] stated that application of multivariate statistical techniques help in the interpretation of complex data matrix and better understanding of water quality and its management. Hence, the study applied Cluster Analysis (CA), Discriminant Analysis (DA) and Principle component Analysis (PCA) to evaluate the variations in the water quality of Mumbai coast; and was able to give meaningful and interesting interpretations to the information obtained.

Further, [11] conducted a sensitivity analysis of water quality parameters using Monte-Carlo analysis (MCA). Also, [12] provides basis for parameter selection during model optimization.

Besides, an oeuvre of studies provide aggiornamentos on water body pollution and water quality management see for example: [13] for U-Tapao River basin Thailand, [14] for River Klang in Malaysia, [15] for Pisuerga River, Spain, [16] for Feitsui Reservoir watershed in northern Taiwan, [17] for urban sub-watersheds in Patiala city in India, [18] for coastal watershed of Southeast China and [19] for Fuji river basin in Japan.

Moreover, the stalking of the level of dissolved oxygen in a body of water based on the observed variability in the level of pollutant in effluent loading appears to fully describe the mechanics of fluctuation of assimilative capacity of a river subject to effluent loading. Some of the foregoing statistical techniques categorically corroborate our approach of deployment of regression model and the accompanying resampling techniques in the nature of bootstrapping to show how adjustment of regression parameters can help to explain why assimilative capacity of a body of water fluctuates. The procedure therefore tries to bring out the level of significance of bootstrapping as applied therein. The procedure also brings out the important role which R^2 , i.e. the multivariate coefficient of determination, plays in explaining to what extent the multivariate linear regression is able to account for the perceived variability or fluctuation of assimilative capacity given some changes in the level of pollutants in effluent loading. It is evident from the foregoing analysis of the previous works that the balance of literature is deficient on the use of regression analysis and bootstrapping to undertake sensitivity analysis to

do tricking in the prediction of dissolved oxygen from sample realizations.

The aim of this study is to use multivariate linear regression to show that geometric projection is a lucid means of identifying unit changes in pollutant variables by which we can further use bootstrapping technique to establish confidence intervals that facilitate tricking of the model parameters with a view to bringing out the necessary adjustments required for the optimization of DO prediction.

The current study offers the use of multivariate linear regression and bootstrapping statistical analysis to investigate the degradation of assimilative capacity of water body subject to effluent loading. The theoretical formulation underpinning the statistical application is provided and it makes the application to be better understood by readers.

2. METHODOLOGY

2.1 Data Collection

The accompanying data depicted in Table 1 were readings obtained from Ikpoba River from three point sources. At each point upstream and downstream, samples of polluted water were obtained 20m upstream and downstream respectively. Again, samples were collected 200m downstream of the third point source. These were taken over three seasons- namely wet, dry, and harmatan season. These samples were analysed in a chemical laboratory according to Table 1.

The observations of the sample pollutants were organized according to the format of Table 2 which is compatible with hat matrix application. The data were fitted into the accompanying Multivariate Linear Regression model.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_7 \beta_7 \quad (1)$$

Where $[x_1, x_2, \dots, x_n]$ are the observations shown in Tables 1 and 2. The problem was solved by both manual and MATLAB software.

2.2 Data Analysis

The Multivariate linear regression model was applied using sample realizations in order to develop a model that can predict the DO, which is the prime determinant of the assimilative capacity of the water body.

The various pollution parameters were treated as predictor variable x_i while the resultant DO was treated as the response variable, y . Also, β_j , $j = 1, 2, \dots, 7$, represents the propensity of each sensitivity index to, in the first instance, adsorb oxygen and subsequently to desorb same in order to meet both chemical and biological oxygen demand and thereby causing diminution of DO in the body of water. With this estimate, the level of DO in the river can be predicted howbeit as point estimates.

On the other hand, by bootstrapping technique of [20], there were 100 times resampling with replacement from the bowl. With this, for each resampling, the tendency of the pollutants to deplete the DO (β) was calculated; and the overall values of β_j determined by means of scree plots see Figure 3 (a-h). The upper and lower control limits (the confidence intervals) of the resultant output of y_i were determined by tricking, involving the substitution of the parameters in order to get the extreme values of y . By substituting the extreme values of each of the model parameters, the sensitivity analysis on the value of y was seen to be evaluated. This defines the "bandwidth" of the pollutants depletive effects on DO in the polluted water body.

Previous to this stage, the multivariate coefficient of determination R^2 was evaluated in order to ascertain the model adequacy. In other words, R^2 which is given by $R^2 = \left[1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}\right]$ ascertains to what extent the model is able to account for the perceived variabilities in the value of response variable y , which is the dissolved oxygen (DO).

Table 1: Data Matrix of Ikpoba River Quality Parameters

	PO ₄	NO ₃	Cd	Cu	Fe	Pb	Turbidity	Temp.	Feacal col.	Ph	DO	BOD	COD
P1 (ups)	0.317	0.491	0.346	0.259	1.552	0.194	2.5	26.467	16.333	7.443	8	3.267	25
P1 (dns)	0.489	0.724	0.316	0.191	1.425	0.137	4.333	25.066	26.667	7.827	7.567	5.767	39.667
P2 (ups)	0.246	0.629	0.641	0.245	1.467	0.155	2.367	27.033	20.667	7.457	6.8	0.7	28.4
P2 (dns)	0.334	0.707	0.732	0.267	2.737	0.126	4.233	27.467	25	7.27	6.433	4.1	28.133
P3 (ups)	0.432	0.499	0.4	0.137	2.097	0.135	2.633	27.8	12	7.41	7	1.267	27.933
P3 (dns)	0.868	0.933	0.77	0.503	5.062	0.717	30	27.433	126.33	6.16	0.267	16.467	277.633
200M dns	0.133	0.487	0.399	0.049	1.108	0.039	2.933	26.7	13.333	7.52	6.667	1.733	12.933

3. THEORETICAL FORMULATION

Linear regression, in a geometrical sense, represents projection of vector of response variable, y_i , onto the space defined by $n \times m$ vector of predictor variable x . Geometrically, we can do the following representation.

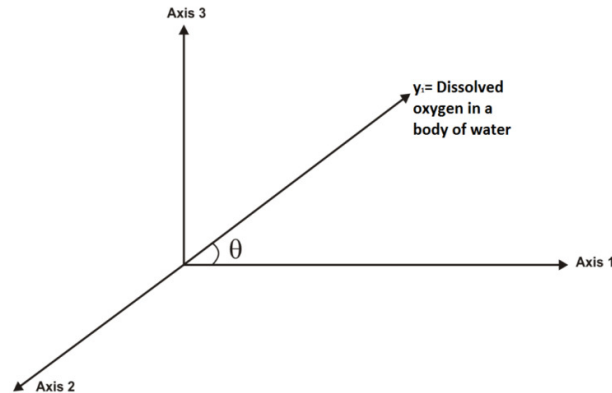


Figure. 1: A 3D representation of (DO)

The plane formed by axes (1) and (2) constitutes the column space on which y is projected in order to form a linear combination of the pollutant vectors that spanned the space defined by axis 1 and axis 2.

Effluents are discharged into the river. The effluent mixes in the water and takes up dissolved oxygen in the water body by both adsorption and desorption (elution) This process repeats itself from time to time. Our goal in modelling is to replicate this natural process in some way. To achieve the goal, a realization is captured to obtain estimates of β which is b . In

other words, b is a ballpark of β . In point of fact, β is the potential or valance or adsorptive capacity which each pollutant has for oxygen adsorption and when the resultant oxygen adsorption and desorption from DO is done, the balance of DO in the water body is y , and the estimate (predicted value) is \hat{y} .

$$\hat{y} = b_0 + bx + \varepsilon \quad (2)$$

In the natural process, we can represents these as column vectors.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \dots, x = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$$

Notice that x is a matrix that is considered tall and skinny; $n \gg m$. The sole purpose of this research is to build a machine that can replicate the perceived natural process (model building). The realization is a sample obtained from the polluted river.

In the natural setting, the model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_5 x_5 + \beta_7 \beta_7 + \varepsilon_i \quad (3)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \dots + \beta_m x_{1m} + \varepsilon_1 \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \dots + \beta_m x_{2m} + \varepsilon_2 \\ \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \beta_3 x_{33} + \dots + \beta_m x_{33} + \varepsilon_3 \\ \vdots \\ \beta_0 + \beta_n x_{n1} + \beta_n x_{n2} + \beta_n x_{n3} + \dots + \beta_n x_{nm} + \varepsilon_m \end{bmatrix}$$

Table 2: Realizations, $X = f(S_{ij}/y_i)$

S/N	Observations of Sample Pollutants							Resultant Depleted DO
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	
1	S ₁₁	S ₁₂	S ₁₃	S ₁₄	S ₁₅	S ₁₆	S ₁₇	y ₁
2	S ₂₁	S ₂₂	S ₂₃	S ₂₄	S ₂₅	S ₂₆	S ₂₇	y ₂
3	S ₃₁	S ₃₂	S ₃₃	S ₃₄	S ₃₅	S ₃₆	S ₃₇	y ₃
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	S _{n1}	S _{n2}	S _{n3}	S _{n4}	S _{n5}	S _{n6}	S _{n7}	y _n

$$\varepsilon \approx N(0, \sigma^2)$$

This model is assumed, to a good extent, to mimic the natural process of oxygen diminution in the body of water as a result of adsorption and desorption of DO.

And with sample realizations we want to usually have $\frac{n}{m} > 100$ and in any case where $\frac{n}{m} \leq 5$, we are bothered that we do not have sufficient observations for estimating the number of parameters in β .

Thus the multiple linear regression equation, in short hand, compact or matrix form, is $y = \begin{matrix} x \\ nxm \end{matrix} \begin{matrix} \beta + \varepsilon_{nx1} \\ mx1 \end{matrix}$ And in expanded linear algebraic form we have :

$$\text{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2_{ij} \sigma; \sigma_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Recall that $b \approx \beta$, implying that b is a ballpark of β . Also β is taken as an estimate.

We shall estimate b using sample realization such that $\hat{y} \approx y_i$. Of course, the model fitting is seen from the point of view of ordinary least squares (OLS).

Let's choose a criterion, Q

$$Q = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (4)$$

where Q is deemed to be the sum of squared deviations or error sum of squares. This quantity, Q , can be seen as a distance or length which reflects the degree or length of departure of estimate, \hat{y} , from the original observation.

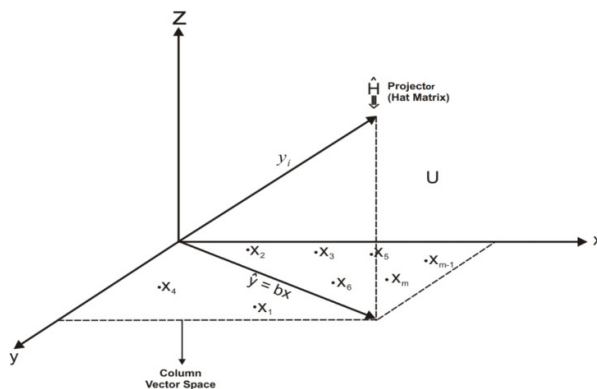


Figure 2: Projection of DO on Column Space by Projector Matrix, \hat{H} .

$$\hat{y} = b_0 + b_1 x_1 + \dots + b_m x_m + \varepsilon_i \quad (5)$$

$$\Rightarrow y - \hat{y} = U \quad (6)$$

U needs to be small so that \hat{y} will be closest to y_i as possible.

Observe that \hat{y} is the projection of y_i on column vector space spanned by $\{x_1 x_2 \dots, x_m\}$

We wish to estimate the values of b that will minimize the value of U

Where

$$\sum U^2 = Q$$

Hence,

$$Q = \sum U^2 = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (7)$$

$$= \sum_{i=1}^n (y_i - x b)^2 \quad (8)$$

Note that equation (7) is in matrix form. Thus $y = xb$ is a matrix difference. Consider:

$$\sum \frac{(y - xb)(y - xb)}{1 \times n \quad 1 \times n} = Q \quad (9)$$

And by matrix operation, this is not conformable. Thus we can use the transpose of one of them to multiply the other since Q is a squared length of vector difference.

Hence we have,

$$Q = \frac{(y - xb)^T}{1 \times n} \frac{(y - xb)}{1 \times n} \quad (10)$$

Equation (11) is now conformable $(1 \times n) \times (n \times 1) = 1 \times 1 = \text{a number}$

In calculus, we know that by equating the differential coefficient to zero, we can obtain the value of b that minimizes Q . Thus we obtain the following sequence

$$\frac{\partial Q}{\partial b_0}, \frac{\partial Q}{\partial b_1}, \frac{\partial Q}{\partial b_2}, \frac{\partial Q}{\partial b_3}, \dots, \frac{\partial Q}{\partial b_m}$$

When each of these expressions are equated to zero, we can after some cumbersome work, obtain the needed values of b that minimize Q . To circumvent this imputed messy calculations, we can work backwards as follows;

Define the hat matrix H

$$\hat{H} = X(X^T X)^{-1} X^T \quad (11)$$

X^T is the transpose of X ; it is the set of observations of pollutants represented as a vector matrix.

Some properties of \hat{H} :

1. The inverse $(X^T X)^{-1}$ exists; in other words, X has full rank
2. \hat{H} is idempotent; hence $\hat{H} \hat{H} = H$
3. Diagonal element of $H = h_{ii}$; it has i^{th} leverage value.

Set $y = \hat{H}y + (1 - \hat{H})y$

We can now develop Q based on the following lemmas:

Lemma 1: $\hat{H}^T = H$

Lemma 2: $(xb)^T = b^T x^T$

Hence; we state as follows:

$$Q = (y - xb)^T (y - xb) \quad (12)$$

$$\begin{aligned} & [\{\hat{H}y + (1 - \hat{H})y\} - xb]^T [\{\hat{H}y + (1 - \hat{H})y\} - xb] \\ \Rightarrow Q &= [\{\hat{H}y - xb\} + (1 - \hat{H})y]^T [\{\hat{H}y - xb\} \\ & \quad + (1 - \hat{H})y] \\ \Rightarrow (\hat{H}y - xb)^T (\hat{H}y - xb) &+ (\hat{H}y - xb)^T (1 - \hat{H})y \\ & \quad + [(1 - \hat{H})y]^T [\hat{H}y - xb] \\ & \quad + [(1 - \hat{H})y]^T [(1 - \hat{H})y] \quad (13) \end{aligned}$$

Notice that the second and third summands vanish as a result of the following manipulation/substitution

$$\begin{aligned} \hat{H} - X(X^T X)^{-1} X^T X \\ (1 - \hat{H})X &= X - \hat{H}X = X - X(X^T X)^{-1} X^T X \\ &= X - X \underbrace{(X^T X)^{-1} (X^T X)}_{=1} = X - X = 0 \quad (14) \end{aligned}$$

Hence we have

$$\begin{aligned} Q &= (\hat{H}y - xb)^T (\hat{H}y - xb) \\ & \quad + [(1 - \hat{H})y]^T [(1 - \hat{H})y] \quad (15) \end{aligned}$$

If Q should be minimized with respect to b , it should be achieved through the first summand (first term of equation 15)

It is noticed that if we set

$$b = (X^T X)^{-1} X^T y \quad (16)$$

Then,

$$\begin{aligned} & (Hy - xb)^T (Hy - xb) \\ &= [Hy - X(X^T X)^{-1} X^T y]^T [Hy - X(X^T X)^{-1} X^T y] \\ &= [X(X^T X)^{-1} X^T y - X(X^T X)^{-1} X^T y]^T [X(X^T X)^{-1} X^T y \\ & \quad - X(X^T X)^{-1} X^T y] \quad (17) \end{aligned}$$

Hence $b = (X^T X)^{-1} X^T y$ is the least squared estimate of β . one cannot agree more that this result is most significant in the sense that it is a direct computation of regression parameters, β , using the format of matrix of realizations $x = f(S_{ij}|y_i)$ depicted in Table 2. It is easily computed with MATLAB software.

3.1 Interpretation of the Model Estimation

β is the estimation of the rate, potential or valence with which the pollutants adsorb oxygen from the body of water. This rate has been estimated from the sample realization to be b . With this estimate of model parameters we can predict the level of dissolved oxygen as

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n + \varepsilon_i \quad (18)$$

Finally, y is the set of observations of the balance of dissolved oxygen level remaining after some quantities had been adsorbed and desorbed by pollutants.

At this juncture, we note that b β are mere point estimates of parameters. We need to use

bootstrapping methodology to estimate the confidence interval, the purpose of which is needful for *model tricking* and adjustments in order to optimize y , the prediction of dissolved oxygen, DO.

4. RESULTS AND DISCUSSION

The results of the study are presented in the following order:

- (i) Regression Analysis
- (ii) Bootstrapping
- (iii) Sensitivity Analysis
- (iv) Check for model Adequacy

We take them seriatim

4.1 Regression Analysis

In line with theory of multiple linear regression (MLR) using hat matrix, H , as matrix projector which elucidated the ordinary least squares (OLS) methodology applied, the regression parameters were computed using both manual and MATLAB software application. The results obtained with MATLAB are as shown in Table 3.

The two results are essentially the same mutatis mutandis, the respective differences being attributed to rounding off errors. Thus confirming that the theory straddling the two approaches are in concordance. Further, conflating both sets of results, it is evident that the manually computed results support the results obtained through MATLAB, confirming the authenticity of the values of the regression parameters. Again, each of these parameters represents the rate at which each variable increases with unit change in its value. The β_0 in particular represents an autonomous value that may or may not have direct practical meaning depending on the situation under consideration. It is a value that provides necessary adjustments to the value of DO and that is why it has been referred to as an autonomous value. It could be seen also as a boundary condition.

4.2 Bootstrapping

Further, we reiterate that the results of the regression parameters are mere point estimates. Thus, we need interval estimates of these parameters. Statistically, this appears difficult because we need a special methodology for generating the means and standard deviations in order to obtain the required confidence interval. In consequence, we had a recourse to the bootstrapping methodology developed by [20]. This method had also been applied in [10] in a study to

analyze the 'Sobol' sensitivity of a complex environmental model. In this study, however, the sensitivity indices were used to rank the contribution and influence of each pollutant to the overall

degradation of the water body. Never the less, this application corroborates the approach adopted by the current study. The results of bootstrapping are shown in Figure 4.

Table 3: MATLAB computed results of Regression analysis

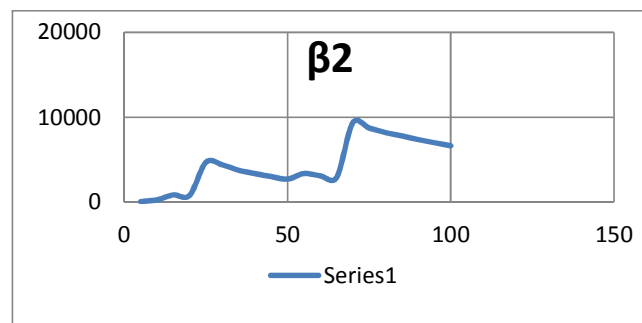
β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
-279.537	80.2411	-131.07	4.0714	3.7552	27.0139	-3.1380	-1.3322

Table 3: Depiction of Confidence Limits across Some Significance Levels

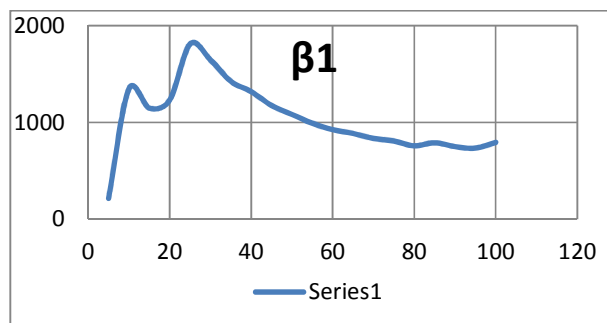
90%	UCL	193.6897	193.0304	192.3905	191.8902	195.5626	184.3446	192.967
	LCL	-177.391	-178.05	-178.69	-179.19	-175.518	-186.736	-178.113
	PNT. EST.	8.149565	7.490211	6.850315	6.350073	10.02246	-1.19555	7.426784
95%	UCL	229.2187	228.5594	227.9195	227.4192	231.0916	219.8736	228.4959
	LCL	-213.579	-213.579	-214.219	-214.719	-211.047	-222.265	-213.642
	PNT. EST.	8.149565	7.490211	6.850315	6.850315	10.02246	-1.19555	7.426784
99%	UCL	299.1487	298.4894	297.8495	297.3493	301.0216	289.8036	298.426
	LCL	-282.85	-283.509	-284.149	-284.649	-280.977	-292.195	-283.572
	PNT. EST.	8.149565	7.490211	6.850315	6.350073	10.02246	-1.19555	7.426784

Gloss: UCL: Upper Control Limit, LCL: Lower Control Limit, PNT. EST.: Point Estimate

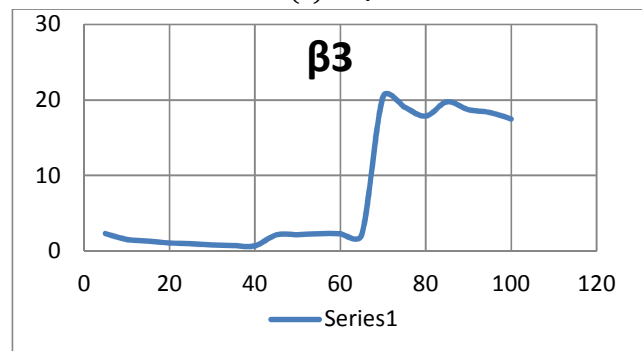
The results of the scree plots are as follows:



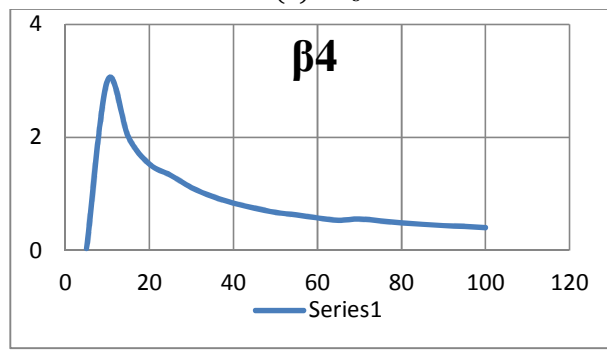
(a) PO₄



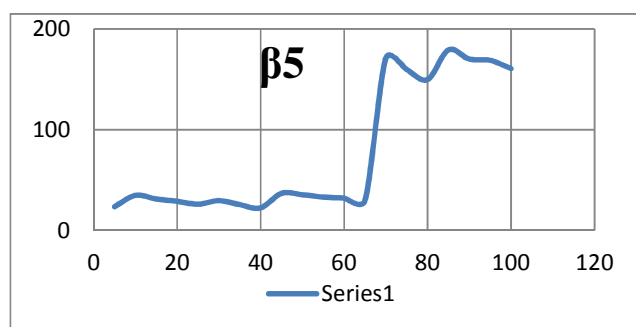
(b) NO₃



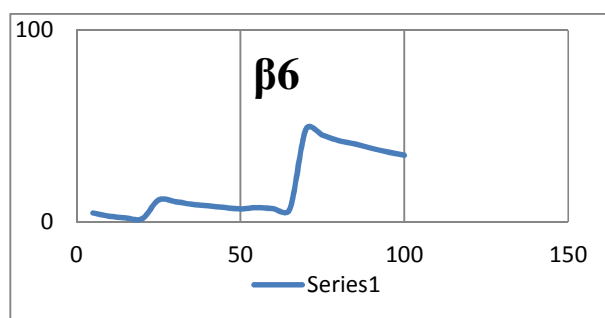
(c) Temp



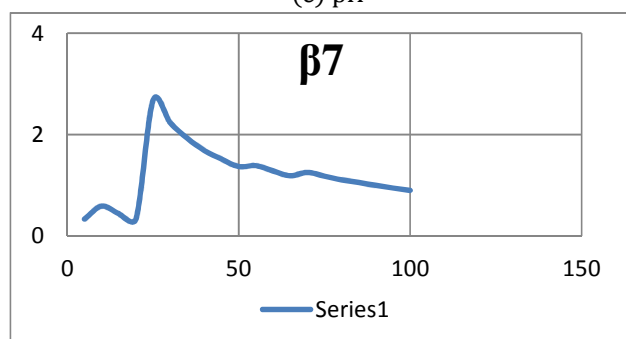
(d) Faecal Coliform



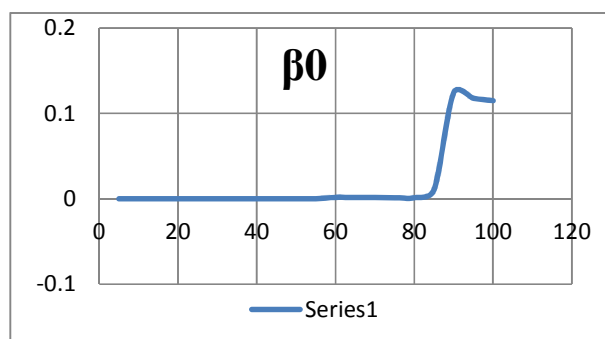
(e) pH



(f) BOD



(g) COD



(h) Autonomous value

The behaviours of β_1, β_4 and β_7 show that their associated variables (PO_4 , Feacal Coliform and COD) are of stable characteristics and therefore, do not cause wild fluctuations in the depletion of DO.

However, we need to watch out for pollutants with the sensitivity indices $\beta_2, \beta_3, \beta_5$ and β_6 (i.e. NO_3 , Temperature, pH and BOD in that sequence). Their values fluctuate significantly and could cause instability in the value of dissolved oxygen, DO. They engender wild fluctuations which lead to high variability in the value of DO and hence they are very sensitive parameters. These parameters account for very severe variations in the value of DO. In other words, they can be said to have higher propensity to adsorb the DO from the body of water by virtue of their erratic characteristics as evidenced in figures 3 (b,c,e and f). β_0 is an autonomous parameter that depends on the boundary condition of pollutant loading. It may or may not have practical interpretations.

4.3 The Confidence Intervals for Sensitivity Analysis

By substituting at once, the values of the upper control limit (UCL) and the lower control limit (LCL) of the sensitivity indices, β_i , each set at a time, into the regression model, the extreme values of DO can be obtained and these extreme values define the limit beyond which the DO cannot exceed. Thus, when the DO values obtained with point estimate (PNT. EST.), which is as an average value, we will then claim that such average values will not fluctuate beyond the

extreme values determined by the UCL and LCL. This is the real essence of the sensitivity analysis because it answers the “what if” question.

The foregoing confidence limits appear to be the pith of the results in the sense that they serve as a tool for model tricking in order to obtain the pattern of fluctuation of the values of DO given the fact that each of the parameters have a range of values within which they fluctuate. Thus by the method of model tricking, it will be possible to eyeball the manner or pattern in which DO deficiency happens as effluents are ejected into the river. This process of model tricking is indeed an optimization technique that provides a means for sensitivity analysis. Thus as the values of the regression parameters fluctuate, so do the value of DO varies to affect the assimilative capacity of the river.

(iv) Check for Model Adequacy

The computed result of the multivariate coefficient of determination, R^2 , show that the value is 0.71 (71%) indicating that a good fit was achieved by the regression model developed.

5. CONCLUSION

This study has ably demonstrated a combined use of multivariate linear regression modelling and bootstrapping as an effective statistical analytics for conducting sensitivity analysis in order to evaluate the robustness of assimilative capacity modelling in a body of water. The philosophy behind sensitivity analysis is a close examination of what amount of changes in the critical values of regression parameters

corresponding to sensitive pollutants, namely NO₃, Temperature, pH and BOD can cause wild fluctuations in the value of DO without affecting the model robustness. The method proposed crosses the stream where it is shallowest. This study has struggled, in a savoir-faire manner, as it were out of a puzzling issue into un hoped-for unfoldment of a vista of knowledge beyond the specific ways of looking at the world. Earlier methods appear to dawdle.

6. REFERENCE

- [1] Igboanugo, A.C and Chiejine, C.M. (2004) Modelling of Far-Field Mixing of Industrial Effluent Plume in Ambient Receiving Water: The Ikpoba River Hub Example Nigerian Journal of Technology (NIJOTECH) Vol. 33. No. 2, April 2014, pp. 207 – 214.
- [2] Igboanugo A.C., Chiejine, C.M. (2011) A Mathematical Analysis of Brewery Effluent Distribution in Ikpoba River in Benin City, Nigeria. *ARPJ Journal of Engineering and Applied Science* Vol. 6, Number 9 September 2011.
- [3] Igboanugo, A.C and Chiejine, C.M. (2012). Pollution Survey of Ikpoba River, Benin City, Nigeria. *Journal of Emerging Trends in Engineering and Applied Sciences* 3 (3): pp 567-571.
- [4] Igboanugo, A.C., Ezemonye, L. I .N.and Chiejine, C.M., (2013). Influence of Effluent Discharge and Runoffs into Ikpoba River on its Water Quality. *Nigerian Journal of Technology* 32(2) July 2013. pp274-303
- [5] Chiejine, C.M., Igboanugo, A.C. and Ezemonye, L.I.N (2015) Modelling Effluent Assimilative Capacity of Ikpoba River, Benin City, Nigeria *Nigerian Journal of Technology Vol.34, No.1*, pp. 133-141.
- [6] Reghunath, R., Murthy, T.R.S., Reghavan, B.R. (2002) The utility of Multivariate statistical techniques in Hydro geochemical studies: an example from Karnataka, *Indian water Research* 36:2437-2442.
- [7] Semeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsas, D. Anthemidis, A., Sofoniou, M. and Kouimtzis, T. (2003) Assessment of the surface water quality in northern Greece. *Water research*, 37: pp4119-4124.
- [8] Wu, B., Zhao, D.Y., Zhang X.X. and Cheng, S. (2009) Multivariate statistical study of organic pollutants in Nanjing reaches of Yangtze River. *Journal of Hazardous materials*. 169, pp.1093-1098.
- [9] Krishna, A.K., Satyanarayanan, M. and Govil P.K. (2009). Assessment of Heavy metal pollution in water using multivariate statistical techniques in an Industrial area. A case study from Patancheru, Medak district. Andhra Pradesh, Indian. *Journal of Hazardous Material*. 167: pp. 366-373.
- [10] Nossent, J, Elsen, P, and Bauwens, N, (2011) Sobol sensitivity analysis of a complex environmental model. *Environmental Modelling and Software* 26 (2011) pp1515-1525.
- [11] Gupta, I., Dhage, S. and Kumar R. (2009) Study of variation in water quality of Mumbai Coast through multivariate analysis techniques. *Indian Journal of Marine science*. 38(2) pp170-177.
- [12] Krause, P., Bende-Michl, I., Fink, M., Helmschrot, J., Kraliseh, S. and Künne, A. (2009) parameter sensitivity analysis of the JAMS/J2000-S Model to improve water and Nutrient transport process simulation – a case study for the Duck Catchment in Tasmania 18th world IMACS/MODSIM congress, Cairns, Australia pp13-17 July 2009. <http://mssanz.org.modsimoa> . [
- [13] Gyawali, S., Techato, K., Yuangyai, C. and Monprapussan, S. (2012) Evaluation of surface water quality using multivariate statistical techniques: a casestudy of U-tapao river basin, Thailand. *KMITL. Science Technology Journal* 12(1) pp7-19.
- [14] Nasir, M.F.M., Samsudain, M.S., Mohamed, I., Awaluddin, M.R.A., Mansor, M. A., Juahir, H and Ramli, N. (2011): River water quality modelling using combined principle component analysis (PCA) and multiple linear regression (MLR): a case study of at Klang River, Malaysia. *World Applied Science Journal. 14 (exploring pathway to sustainable environmental Issues)* pp 73-82.
- [15] Helena, B., Pardo, R., Vega. M., Barrado E., Fernandez, J.M., Fernandez, L. (2000) temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principle component analysis. *Water Research* 34:807-816.
- [16] Wu, E.M and Kuo S.L (2012) Applying a multivariate statistical analysis model to evaluate the water quality of a watershed. *Water environment Research* 84 (12) pp. 2015-2085.
- [17] Arora, A.S and Reddy, A.S. (2013) Development of multiple linear regression models for predicting the storm water quality of urban sub-watersheds. *Bulletin of environmental contamination and toxicology* DO 1.1007/500128-013-1160-y.
- [18] Huang, J., Huang, Y. and Zhang. Z (2014) Couple effects of natural and anthropogenic controls on seasonal and spatial variation of river water quality during base flow in a coastal watershed of southeast China. *Plos one* 9 (3) pp1-19 .
- [19] Shrestha, S. and Kalama, F, (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin Japan. *Environmental Modelling and Software* 22 pp. 464-475.
- [20] Efron, B. and Tibishirani, R (1993) *An Introduction to Bootstrap*. Chapman Hill. New York, USA